International Conference on Computational Science, ICCS 2010

Small World models for social network algorithms testing

Igor Kanovsky*
Max Stern College of Emek Yezreel, Emek Yezreel, 19300, Israel
University of Haifa, Mt. Carmel, Haifa, 31905, Haifa, Israel

Abstract

Social networks have small-world property, hierarchical community structure, and some other properties. This paper proposes models of networks with these properties and algorithm for community structure recognition. The models are useable for testing effectiveness and efficiency of different algorithms for social network analysis.

© 2010 Published by Elsevier B.V.

Keywords: social network, Small World, communities structure, complex networks, clustering, algorithms benchmark.

1. Introduction

Social networks are now in the center of Web algorithms development efforts because of their central role in contemporary Web development. Main tasks are related to community structure recognition and understanding of the social role of different actors by link structure [1]. To test these algorithms effectiveness (how precise is the answer set to the information need) and efficiency (in time and memory usage), researchers need models of social networks which capture main statistical properties of real world networks with well defined community structure. In the model case as opposed to a real network, the correct result of an algorithm is known, and the network size may vary widely, so different algorithms may be tested and compared in different conditions.

The list of main social network properties includes hierarchical community structure [2], Small World property [3], power law distribution of nodes degree [3], self-similarity [4]. For complex networks the models proposed explain power law nature of degree distribution. The most basic is Barabasi-Albert model of scale free network [5]. Other models presented in [3]. But these models have no communities, so they are not useable in the case of social networks. Recently some models were proposed with community structure [6-8], but these models have no small-world property.

In this paper a set of models for social networks is proposed. The models are based on Small World graph of Watts and Strogats [9], which is the simplest model for small-world. By simulation of some link redirection processes, the models with community structure and small-world properties were generated. In addition, algorithm

* Corresponding author. Tel.: +972-4-6423-493; fax: +972-4-6423-517.
E-mail address: igork@yvc.ac.il.
for community recognition in social networks is proposed, which differ from others by networks small-world nature utilizing.

2. Extended Watts and Strogatz (EWS) model

In a social network two nodes linked to the same node most probably linked too. To describe this phenomenon clustering coefficient was introduced [9]. Let a node neighbors \( N_i \) be a set of nodes linked to the node number \( i \). Clustering coefficient (CC) of node \( i \) is a ratio of number of links in \( N_i \) to maximum number of links in a graph with size \( |N_i| \). CC of a network is the clustering coefficient of nodes averaged over network. The small-world is a network with clustering coefficient significantly bigger than one in a random graph with the same size, but having a small average distance between nodes, approximately the same as that of a random graph.

The simplest model of small-world was proposed by Watts and Strogatz [9] and is called Small World graph. It is generated starting with a regular grid and redirecting part \( p \) of the links randomly. The grid has each node linked to its \( z \) neighbors (usually the grid is depicted as a ring). It has a big CC but also a big average distance. If \( p \) is small (~0.05..0.3), the graph has a small average distance but the CC is big. This is Small World graph. If \( p \) is close to 1, the graph becomes a random one.

However this graph has no community structure. From the link topology point of view, community is a sub graph which has a bigger density of inner community links than the density of between communities links [6].

Let us consider a set of \( M \) grids. By redirecting the \( p_{in} \) fraction of links uniformly at random inside the grid and the \( p_{out} \) fraction of links between the grids, the extended Watts and Strogats (EWS) model is created. There are exactly \( M \) communities in the model, the communities disappear with \( p_{out} \) increasing. The small-world property of the model disappears with \( p_{out} \) or \( p_{in} \) increasing. The model for \( M=3 \) is depicted on Fig.1.

To simulate other social network properties, it is useful to create EWS graph in reverse order, starting with random graph \( G_R \). Inversed to randomization, operation \( I(M) \) on graph \( G_R \) starts with nodes labeling by \( M \) different community labels. With probability 1- \( p_{in} \cdot p_{out} \) each link is redirected to be a part of greed pattern inside a community. With probability \( p_{in} \) a link is redirected to by an inner community random link and with probability \( p_{out} \) to be intercommunity random link. The result of \( I(M)G_R \) is a EWS graph. The \( I(M) \) operation preserves graph size, so it is defined by \( G_R \) size.

3. Hierarchical communities (HC) model

In social networks there exists a hierarchy of communities such that each community consists of sub communities; a sub community has its sub communities, and so on till some level \( h \). To generate a structured graph, lets introduce community collapsing operator \( C \). Graph \( C \cdot G \) is obtained from \( G \) by communities collapsing to a single node. If a graph \( G \) consists of \( M \) communities, graph \( C \cdot G \) has \( M \) nodes linked by the between community links of \( G \) (duplicated links in \( C \cdot G \) graph are redirected by following \( I \)-operation). Collapsed graph may be restored to previous form by \( C^{-1} \) operator. For this purpose the inner structure of communities is preserved for each node of collapsed graph, and, between-communities links are assigned to the restored community nodes randomly.

Let \( \{M_1, M_2,.., M_h\} \) be a set integers equal to of number of community on each level of hierarchy. To generate a small-world network with desired community hierarchy \( G_{HC} \) it is enough to do:

\[
G_{HC}=C^{h+1}I(M_1) \cdot \left[C \cdot I(M_2)\right] \cdot \cdots \cdot \left[C \cdot I(M_h)\right]G_R. \tag{1}
\]

An obtained graph is self-similar if \( M_{k+1}=\mu \cdot M_k \), where \( \mu > 1 \) is the average branching number in a hierarchical tree.
4. Communities recognition in small-world networks

Because social network is a small-world, it is interesting to utilize this property for community recognition. The proposed method invokes link weighting based on the link participants in local links correlation. Inter-community links are perceived as not correlated or weak correlated.

We define that the weight of a link in network is proportional to the number of common neighbours in its neighbourhood. For a link connecting nodes $v_1, v_2$ we define the link weight $\beta$ as:

$$\beta(v_1, v_2) = \frac{|N(v_1) \cap N(v_2)|}{|N(v_1) \cup N(v_2)|}$$

(2)

where $N(v)$ is the neighborhood of the node $v$. Observe that $0 \leq \beta \leq 1$.

Two adjacent nodes $v_1, v_2$ belong to the same community if the weight of their connecting link is bigger than some threshold value, which is a parameter of the algorithm,

$$\beta(v_1, v_2) > \alpha,$$

where $\alpha$ is the level of the community separation for a graph. Notice that non-adjacent nodes belong to the same community if there exists a path $L$ connecting $v_1$ to $v_2$ such that each link in $L$ has a bigger weight than the threshold value. The simplest way for community recognition is to remove all the links with weight below $\alpha$. But we propose another approach, which we consider as more applicable.

Iterated Community Recognition Algorithm (ICRA):

Input: graph $G = (V, E)$, level of community separation $\alpha$;

Output: communities $\{V_1, V_2, ..., V_k\}$;

Start:

\[ i=0 \]

\[ Loop \ while \ V \ is \ not \ empty \]

\[ i++ \]

\[ Find \ an \ arbitrary \ community \ A \ in \ the \ graph \ G[V] \ induced \ by \ V. \]

\[ V_i = A; \quad V = V - A; \]

\[ k = i \]

End.

Algorithm for finding an arbitrary community:

Input: graph $G = (V, E)$;

Output: a community $A \subset G$.

Start:

Put arbitrary node $v \in V$ into queue $Q$;

Loop while $Q$ is not empty

\[ get \ node \ u \ from \ Q; \ add \ u \ to \ A; \]

Loop for each $w \in N(u), \ w \notin C, \ Q$

\[ If \ \beta(u, w) > \alpha \ put \ w \ into \ Q \ End \ If \]

End.

The average computational complexity of ICRA is $z \cdot 2 |E|$, where $z$ is average nodes degree.

The method was tested in EWS and HC models. For different model parameters the fraction of nodes $p_{correct}$ which were correctly classified by the method as belonging to community was calculated by simulation. The $p_{correct}$ value is near 100% for $p_{out}$ or $p_{in}$ less than some critical value $p_c$ and it sharply decreases with ongoing randomization (see Fig. 2). The $p_c$ value is near the same for $p_{out}$ and $p_{in}$, and it increases from approximately 25% to 70% with increasing $z$ from 6 to 16. Other parameters of the model show a weak effect on $p_c$. 
It is important to notice that $p_c$ is the critical value for the clustering coefficient also. Thus the simulations show the strong ability of the method for community recognition in the case of small-world network only. Even if there are communities which may be recognized by other methods, absence of small-world property makes ICRA inapplicable.

5. Conclusion

An algorithm for generating synthetic network with close to real social network properties was developed and tested. It may be used for farther researches both as a base for practical algorithms benchmarks and as a model for social processes simulations. The models proposed in this paper have significant advantage in comparison to previous [6,7], because these are small-world networks with community structure.

The main disadvantages of the model are an absence of a power law distribution (PLD) of nodes degree and not exact simulation of the community size distribution. PLD may be simulated by redirecting random part of links with probability proportional to the degree of nodes (in contrast to absolutely random in the paper). It was shown [5] that this is enough to achieve a PLD. A variety of community sizes may be easily achieved by labeling the required number of nodes as community member. But these additions will increase the model's complexity and number of model parameters, which is better to avoid. In the proposed model there are only few parameters: network size, $p_{out}$, $p_{in}$ and number of communities on each level of hierarchy.

An algorithm for community recognition in social networks was proposed. It is interesting to test different algorithms for community recognition (see [6,8,10]) in framework of these models, which will be done soon.

6. References